BIOIN 401 Final Report
Bryland Schoneck
1473495
April 27, 2021

Supplementary Data Table 1

O - Phylum	Count	O - Class	Count	O - Order	Count	O - Family	Count	O - Genus	Count
Firmicutes	927	Clostridia	825	Clostridiales	817	Coriobacteriaceae	402	Collinsella	346
Actinobacteria	424	Actinobacteria	424	Coriobacteriales	402	Ruminococcaceae	185	Clostridium	219
Bacteroidetes	280	Bacteroidia	274	Bacteroidales	274	Peptostreptococcaceae	214	Prevotella	86
Tenericutes	73	Negativicutes	83	Selenomonadales	83	Lachnospiraceae	118	Bacteroides	43
Proteobacteria	84	Mollicutes	73	Erysipelotrichales	36	Prevotellaceae	92	Ruminococcus	66
Other	44	Other	149	Other	164	Other	437	Other	337
NA	121	NA	124	NA	176	NA	504	NA	855
P - Phylum	Count	P - Class	Count	P - Order	Count	P - Family	Count	P - Genus	Count
Firmicutes	931	Clostridia	824	Clostridiales	815	Coriobacteriaceae	402	Collinsella	346
Actinobacteria	423	Actinobacteria	423	Coriobacteriales	402	Ruminococcaceae	193	Clostridium	143
Bacteroidetes	280	Bacteroidia	274	Bacteroidales	274	Peptostreptococcaceae	144	Prevotella	86
Tenericutes	138	Negativicutes	83	Selenomonadales	83	Lachnospiraceae	133	Bacteroides	50
Proteobacteria	84	Mollicutes	73	Erysipelotrichales	36	Prevotellaceae	94	Ruminococcus	49
Other	53	Other	151	Other	165	Other	486	Other	505
NA	44	NA	124	NA	177	NA	500	NA	773

Results from Check and DIAMOND combined together. O represents Our results where P stands for the Paper's results.

Supplementary Data Table 2a

Paper Full Results	
UMGS179	3517
UMGS183	2770
UMGS84	2592
UMGS264	2009
UMGS470	1973
UMGS2067	1868
UMGS90	1855
UMGS86	1674
UMGS125	1656
UMGS463	1606
UMGS39	1583
UMGS759	1564
UMGS1710	1538
UMGS365	1395
UMGS132	1363
UMGS184	1352
UMGS537	1341
UMGS137	1298
UMGS762	1288
UMGS160	1249

Paper Dataset 1		My Dataset 1	
UMGS179	33	UMGS179	33
UMGS183	18	UMGS183	25
UMGS84	18	UMGS84	18
UMGS264	15	UMGS264	14
UMGS470	10	UMGS470	10
UMGS2067	4	UMGS2067	21
UMGS90	3	UMGS90	6
UMGS86	5	UMGS86	5
UMGS125	13	UMGS125	13
UMGS463	30	UMGS463	30
UMGS39	12	UMGS39	12
UMGS759	5	UMGS759	4
UMGS1710	10	UMGS1710	12
UMGS365	10	UMGS365	11
UMGS132	8	UMGS132	8
UMGS184	3	UMGS184	16
UMGS537	6	UMGS537	6
UMGS137	5	UMGS137	6
UMGS762	16	UMGS762	16
UMGS160	18	UMGS160	18

Paper Dataset 2		My Dataset 2	
UMGS179	44	UMGS179	43
UMGS183	19	UMGS183	23
UMGS84	21	UMGS84	21
UMGS264	21	UMGS264	21
UMGS470	12	UMGS470	12
UMGS2067	3	UMGS2067	21
UMGS90	4	UMGS90	6
UMGS86	8	UMGS86	8
UMGS125	12	UMGS125	11
UMGS463	24	UMGS463	24
UMGS39	7	UMGS39	7
UMGS759	4	UMGS759	4
UMGS1710	8	UMGS1710	6
UMGS365	10	UMGS365	10
UMGS132	6	UMGS132	6
UMGS184	2	UMGS184	19
UMGS537	8	UMGS537	8
UMGS137	3	UMGS137	3
UMGS762	14	UMGS762	14
UMGS160	18	UMGS160	19
Paper Dataset 3		My Dataset 3	
UMGS179	28	UMGS179	28
UMGS183	15	UMGS183	20
UMGS84	10	UMGS84	10
UMGS264	14	UMGS264	14
UMGS470	14	UMGS470	14
UMGS2067	2	UMGS2067	20
UMGS90	5	UMGS90	5
UMGS86	1	UMGS86	1
UMGS125	8	UMGS125	8
UMGS463	17	UMGS463	17
UMGS39	7	UMGS39	7
UMGS759	4	UMGS759	4
UMGS1710	10	UMGS1710	10
UMGS365	8	UMGS365	8
UMGS132	5	UMGS132	5
UMGS184	5	UMGS184	21
UMGS537	13	UMGS537	13
UMGS137	2	UMGS137	3
UMGS762	12	UMGS762	12
UMGS160	18	UMGS160	18

Paper Dataset 4		My Dataset 4	
UMGS179	32	UMGS179	31
UMGS183	13	UMGS183	22
UMGS84	19	UMGS84	19
UMGS264	10	UMGS264	10
UMGS470	14	UMGS470	13
UMGS2067	2	UMGS2067	20
UMGS90	8	UMGS90	9
UMGS86	5	UMGS86	5
UMGS125	9	UMGS125	9
UMGS463	30	UMGS463	29
UMGS39	9	UMGS39	9
UMGS759	4	UMGS759	4
UMGS1710	4	UMGS1710	2
UMGS365	6	UMGS365	6
UMGS132	6	UMGS132	6
UMGS184	4	UMGS184	24
UMGS537	9	UMGS537	8
UMGS137	4	UMGS137	4
UMGS762	16	UMGS762	17
UMGS160	26	UMGS160	26

The counts of how many UMGS were present in the metagenomic subsample, Full paper total of 13,133 and datasets total of 100

Supplementary Data Table 2b

Paper Full Results	
UMGS179	26.78%
UMGS183	21.09%
UMGS84	19.74%
UMGS264	15.30%
UMGS470	15.02%
UMGS2067	14.22%
UMGS90	14.12%
UMGS86	12.75%
UMGS125	12.61%
UMGS463	12.23%
UMGS39	12.05%
UMGS759	11.91%
UMGS1710	11.71%
UMGS365	10.62%
UMGS132	10.38%
UMGS184	10.29%
UMGS537	10.21%
UMGS137	9.88%
UMGS762	9.81%
UMGS160	9.51%

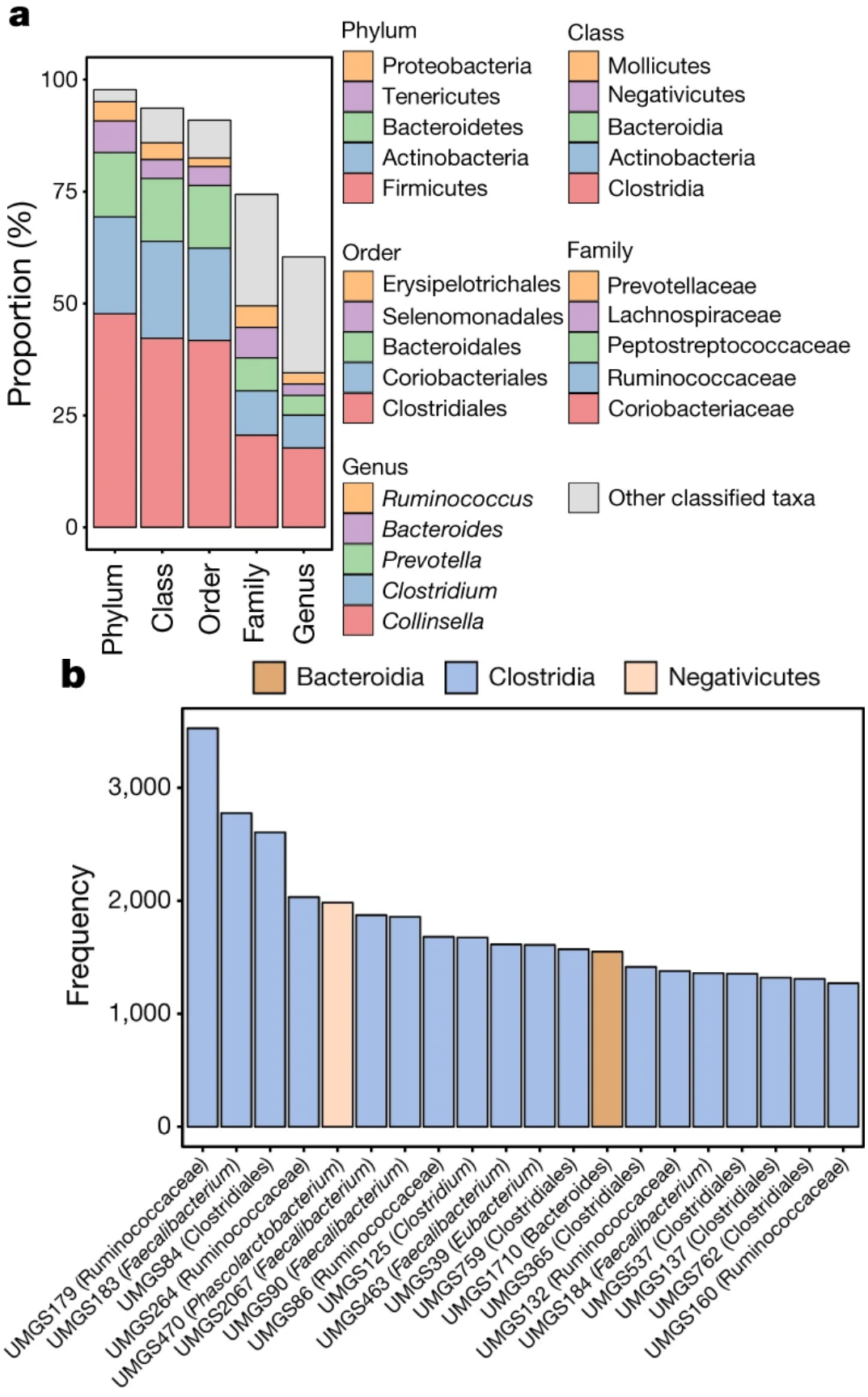
Paper Dataset 1		My Dataset 1	
UMGS179	33.00%	UMGS179	33.00%
UMGS183	18.00%	UMGS183	25.00%
UMGS84	18.00%	UMGS84	18.00%
UMGS264	15.00%	UMGS264	14.00%
UMGS470	10.00%	UMGS470	10.00%
UMGS2067	4.00%	UMGS2067	21.00%
UMGS90	3.00%	UMGS90	6.00%
UMGS86	5.00%	UMGS86	5.00%
UMGS125	13.00%	UMGS125	13.00%
UMGS463	30.00%	UMGS463	30.00%
UMGS39	12.00%	UMGS39	12.00%
UMGS759	5.00%	UMGS759	4.00%
UMGS1710	10.00%	UMGS1710	12.00%
UMGS365	10.00%	UMGS365	11.00%
UMGS132	8.00%	UMGS132	8.00%
UMGS184	3.00%	UMGS184	16.00%
UMGS537	6.00%	UMGS537	6.00%
UMGS137	5.00%	UMGS137	6.00%
UMGS762	16.00%	UMGS762	16.00%
UMGS160	18.00%	UMGS160	18.00%

Paper Dataset 2		My Dataset 2	
UMGS179	44.00%	UMGS179	43.00%
UMGS183	19.00%	UMGS183	23.00%
UMGS84	21.00%	UMGS84	21.00%
UMGS264	21.00%	UMGS264	21.00%
UMGS470	12.00%	UMGS470	12.00%
UMGS2067	3.00%	UMGS2067	21.00%
UMGS90	4.00%	UMGS90	6.00%
UMGS86	8.00%	UMGS86	8.00%
UMGS125	12.00%	UMGS125	11.00%
UMGS463	24.00%	UMGS463	24.00%
UMGS39	7.00%	UMGS39	7.00%
UMGS759	4.00%	UMGS759	4.00%
UMGS1710	8.00%	UMGS1710	6.00%
UMGS365	10.00%	UMGS365	10.00%
UMGS132	6.00%	UMGS132	6.00%
UMGS184	2.00%	UMGS184	19.00%
UMGS537	8.00%	UMGS537	8.00%
UMGS137	3.00%	UMGS137	3.00%
UMGS762	14.00%	UMGS762	14.00%
UMGS160	18.00%	UMGS160	19.00%
Paper Dataset 3		My Dataset 3	
UMGS179	28.00%	UMGS179	28.00%
UMGS183	15.00%	UMGS183	20.00%
UMGS84	10.00%	UMGS84	10.00%
UMGS264	14.00%	UMGS264	14.00%
UMGS470	14.00%	UMGS470	14.00%
UMGS2067	2.00%	UMGS2067	20.00%
UMGS90	5.00%	UMGS90	5.00%
UMGS86	1.00%	UMGS86	1.00%
UMGS125	8.00%	UMGS125	8.00%
UMGS463	17.00%	UMGS463	17.00%
UMGS39	7.00%	UMGS39	7.00%
UMGS759	4.00%	UMGS759	4.00%
UMGS1710	10.00%	UMGS1710	10.00%
UMGS365	8.00%	UMGS365	8.00%
UMGS132	5.00%	UMGS132	5.00%
UMGS184	5.00%	UMGS184	21.00%
UMGS537	13.00%	UMGS537	13.00%
UMGS137	2.00%	UMGS137	3.00%
UMGS762	12.00%	UMGS762	12.00%
UMGS160	18.00%	UMGS160	18.00%

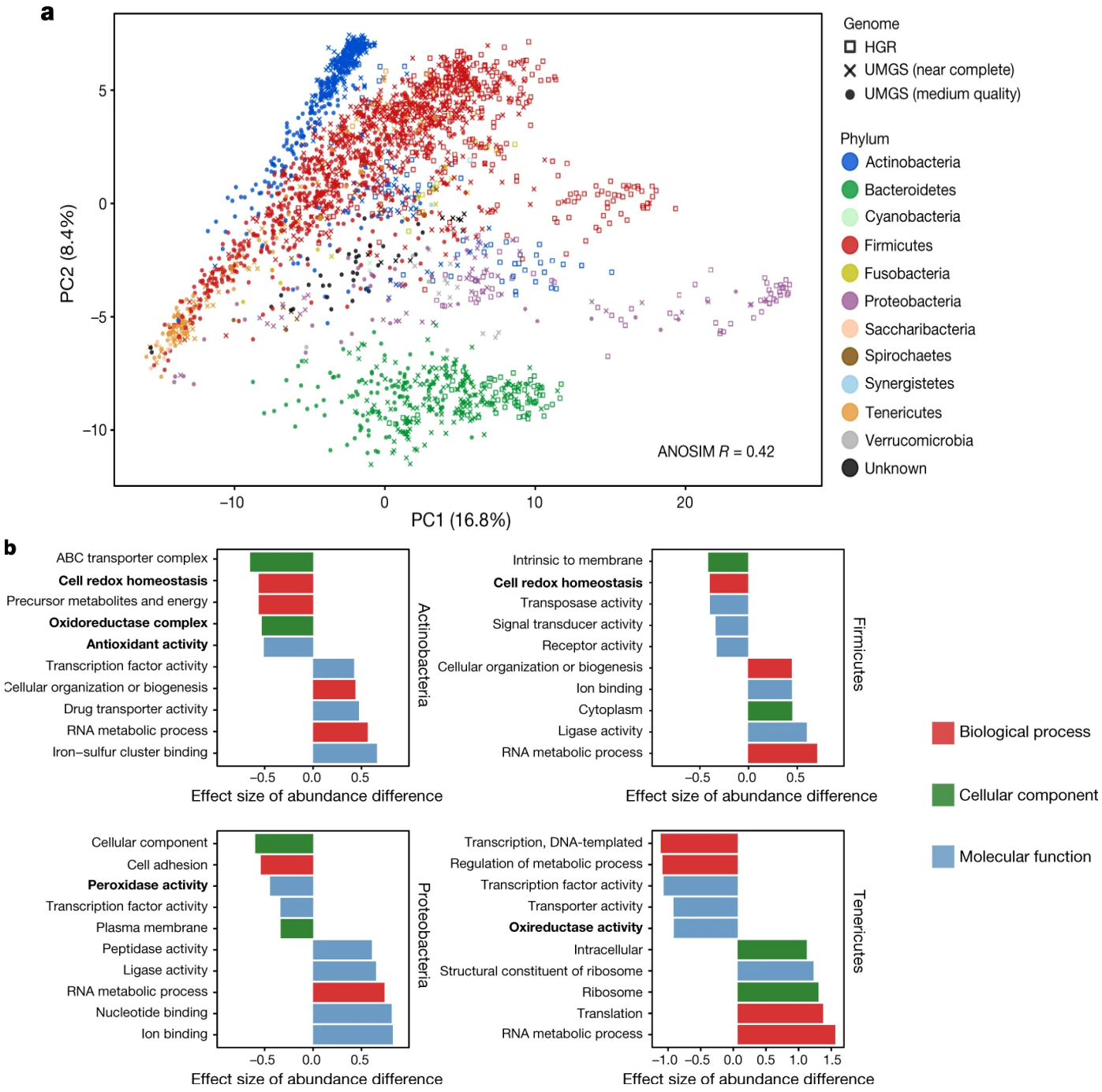
Paper Dataset 4		My Dataset 4	
UMGS179	32.00%	UMGS179	31.00%
UMGS183	13.00%	UMGS183	22.00%
UMGS84	19.00%	UMGS84	19.00%
UMGS264	10.00%	UMGS264	10.00%
UMGS470	14.00%	UMGS470	13.00%
UMGS2067	2.00%	UMGS2067	20.00%
UMGS90	8.00%	UMGS90	9.00%
UMGS86	5.00%	UMGS86	5.00%
UMGS125	9.00%	UMGS125	9.00%
UMGS463	30.00%	UMGS463	29.00%
UMGS39	9.00%	UMGS39	9.00%
UMGS759	4.00%	UMGS759	4.00%
UMGS1710	4.00%	UMGS1710	2.00%
UMGS365	6.00%	UMGS365	6.00%
UMGS132	6.00%	UMGS132	6.00%
UMGS184	4.00%	UMGS184	24.00%
UMGS537	9.00%	UMGS537	8.00%
UMGS137	4.00%	UMGS137	4.00%
UMGS762	16.00%	UMGS762	17.00%
UMGS160	26.00%	UMGS160	26.00%

The proportion of how many UMGS were present in the metagenomic subsample, Full paper total of 13,133 and datasets total of 100

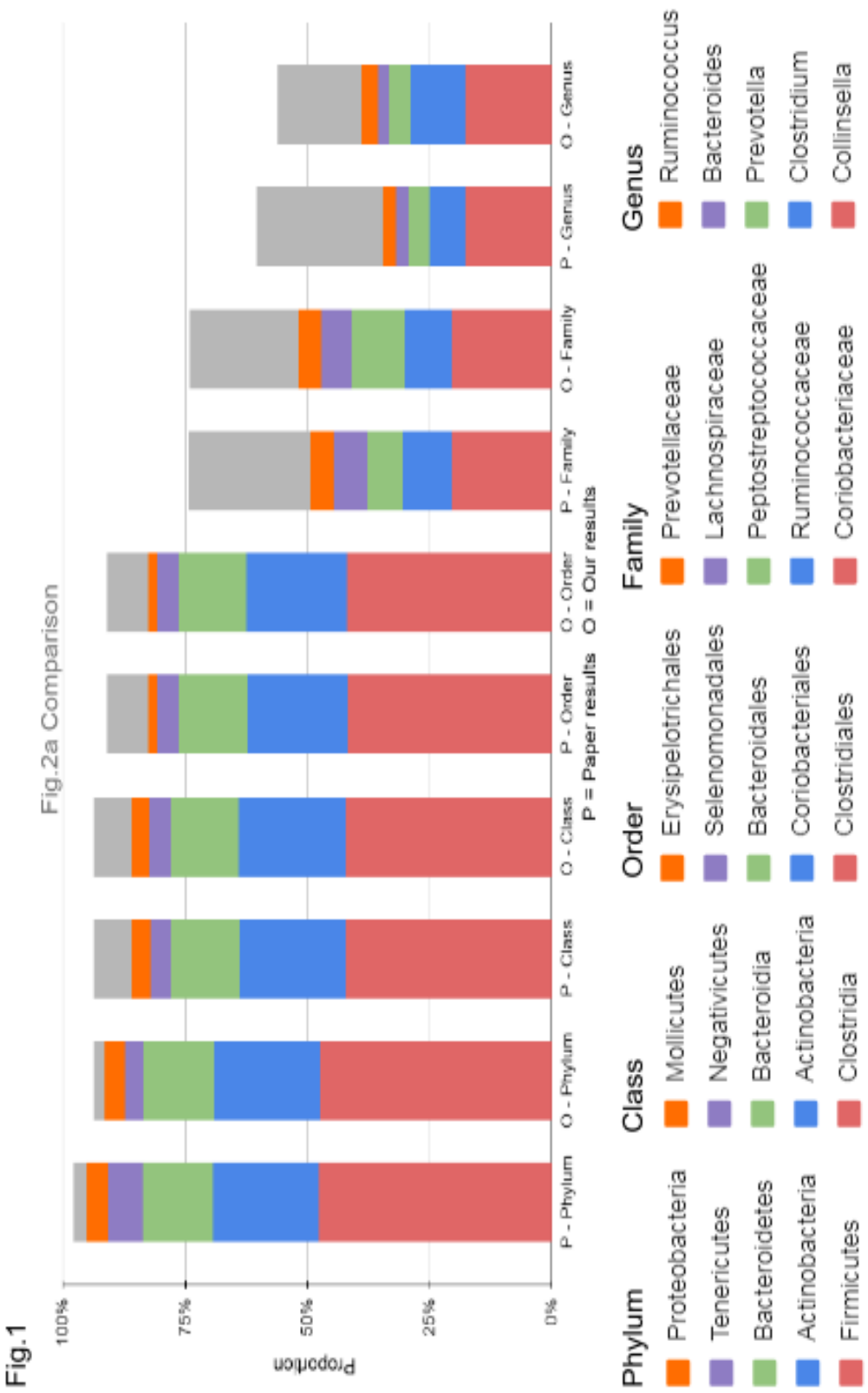
Enlarged Figure 2 from Almeida et al., 2019



Enlarged Figure 5 from Almeida et al., 2019



Enlarged Figure 1



Enlarged Figure 2

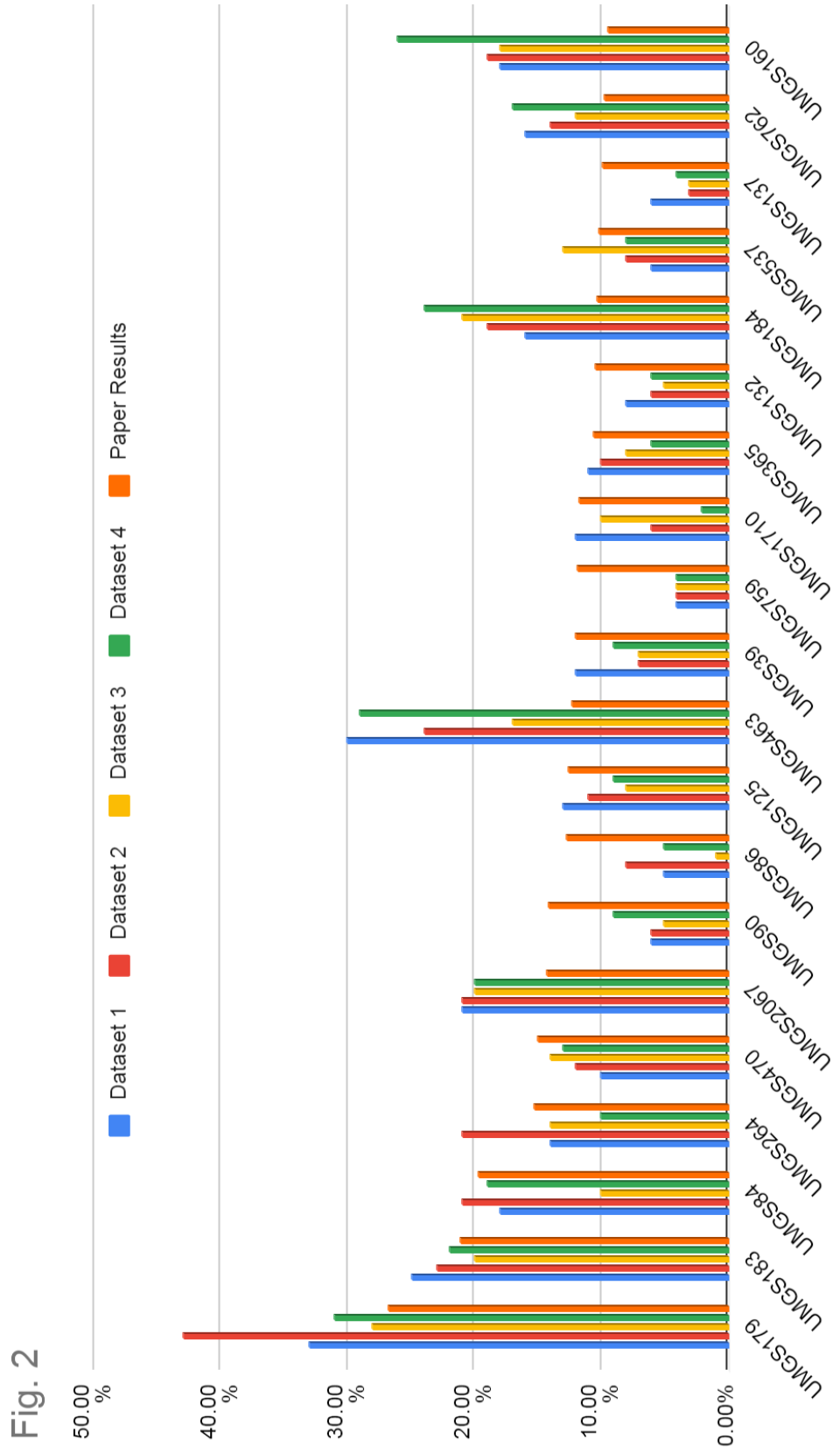
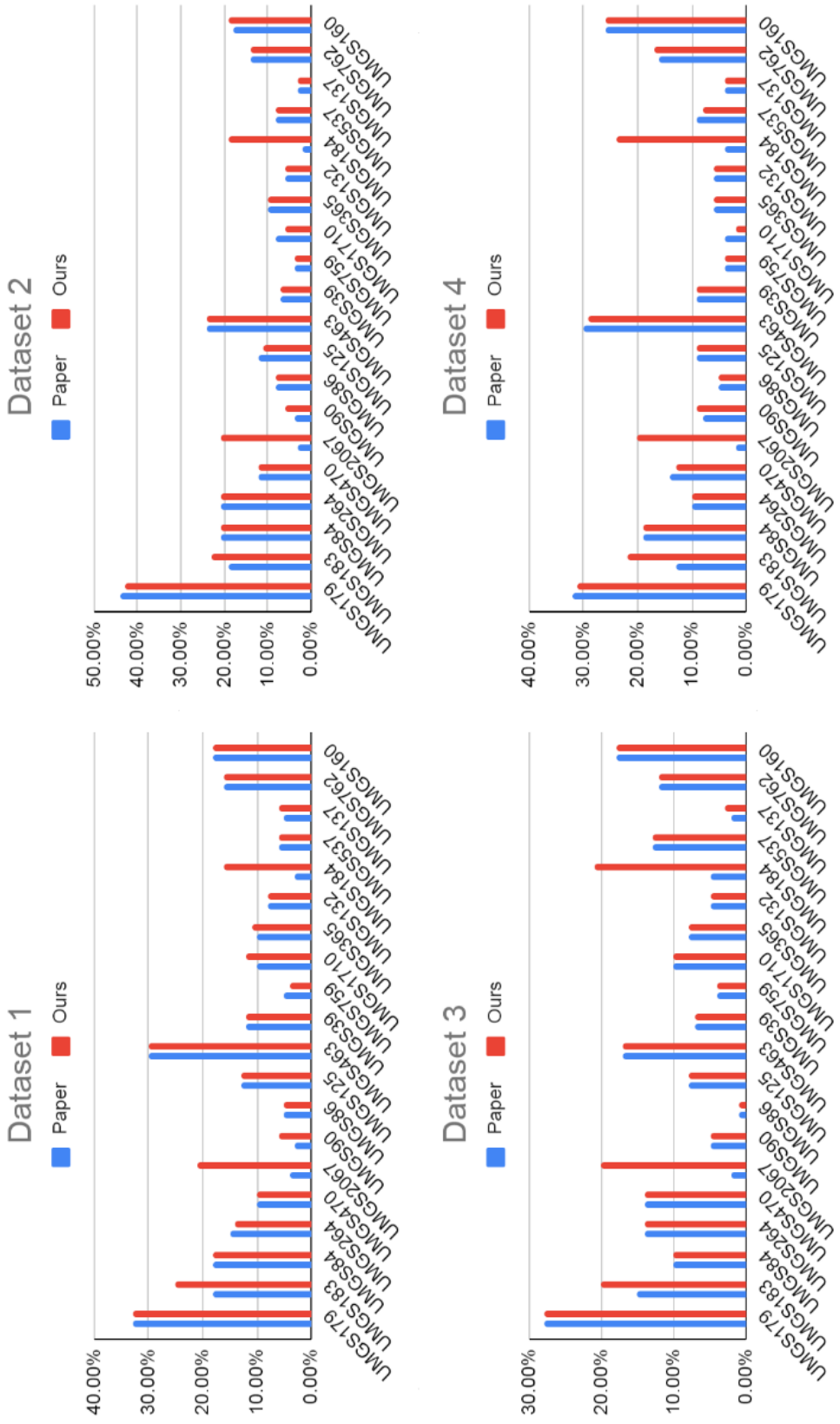


Fig. 2

Enlarged Figure 3



Enlarged Figure 4